DOCUMENT RESUME

ED 193 284 TM 800 604

AUTHOR Byrne, Carolyn D.
TITLE Setting the State Reference Point for the Reading Test for New York State Elementary Schools.
PUB DATE Apr 80
NOTE 20p.: Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, 7-11, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Content Area Reading; Elementary Education; Grade 5; Grade 6; *Minimum Competency Testing; *Reading Achievement; *Reading Comprehension; Reading Diagnosis; *Reading Tests; State School District Relationship; State Standards
IDENTIFIERS *New York; *New York State Basic Competency Test in Reading

ABSTRACT

How the reference point for the Reading Test for New York State Elementary Schools was set is described. A sampling plan was used to define de facto reading standards in grades 5 and 6. Textbooks in English-Language Arts, science, social studies, health, and safety most used in those grades were determined by mailing a questionnaire to the superintendents of 41 randomly selected school districts in New York. Passages from the texts were analyzed using the Bormuth readability formula, and then converted to Degrees of Reading Power (DRP) units. Results indicated that the readability of textbooks was about the same for grades 5 and 6, and was relatively homogeneous across subjects and within subjects for grade 6. Administered at the beginning of grade 6, the standard reference point for the test was based upon the mean readability in DRP units of grades 5 and 6 textbooks combined. Remediation is required for students scoring below the reference point. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED193289

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

New York State's Competency Test in Reading

Setting the State Reference Point for the Reading Test for New York State Elementary Schools

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. Byrne

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Carolyn D. Byrne
New York State Education Department

TM 800604

Paper presented at an AERA symposium "New York State's Competency Tests in Reading and Writing", Boston, Massachusetts, April 11, 1980

Introduction

The Degrees of Reading Power (DRP) which is designed to measure the most difficult prose text a student can read with comprehension is an important part of New York State's Competency Testing Program. The test is administered to all sixth grade students as part of the Pupil Evaluation Program, to selected eighth or ninth graders as part of the Preliminary Competency Testing Program and to certain eleventh or twelfth grade students as part of the Regents Competency Testing Program.

The DRP consists of nonfiction prose passages on randomly selected topics. The passages each contain about 300 words and are arranged in order of difficulty beginning with very easy material and progressing to very difficult material. Test items are formed by the deletion of seven words in each passage, with the student selecting the most appropriate word from the five options provided for each deletion. In order to answer DRP questions correctly, the student must read and comprehend the text pertaining to those items. This represents a clear contrast with some conventional reading tests, where questions can sometimes be answered without reading the pertinent text.

The exact composition of a DRP test depends on the purpose or function of the program for which the test was designed. For example, when DRP methodology is used for the Regents Competency Test, the passages are concentrated around the standard established for high school graduation. When it is used in the Preliminary Competency Test, a broad range of passages is used in order to identify students in need of various types of remedial activity. When DRP methodology is used in the sixth grade Pupil Evaluation Program, the tests are tailored to student ability so that maximum information for placement, needs assessment, and growth can be provided. At the sixth grade there are two forms of the test one designed for those students reading below grade level and the other for those students reading at or above grade level.

A reference point or standard has been established for each of the state tests, the Regents Competency Test in Reading, the Preliminary Competency Test in Reading and the Reading Test for New York State Elementary Schools. In each case the reference point is a standard expressed in terms of readability. For the Regents Competency Test this point represents a sixty-five percent likelihood of success on prose materials typically found in textbooks used in tenth and eleventh grade. For the Preliminary Competency Test it will represent a seventy percent likelihood of success on prose materials typically found in eighth and ninth grade textbooks. On the

Reading Test for New York State Elementary Schools the point represents a seventy percent likelihood of success on prose materials typically found in textbooks used in grades five and six.

The setting of standards for mandated competency requirements is one of the more important and controversial problems confronting the entire educational community, and it has special salience for the measurement specialist. (Linn, p 302).

Although it is of fundamental importance, the problem of setting standards has received very little attention in the measurement literature. A standard often is assumed to have been established and then is used as the starting point to develop techniques for determining whether an examinee should be classified above or below the standard. With some exceptions (Millman, 1973; Meskauskas, 1976) little attention has been directed to the question of where the standards come from, who establishes them, and what procedures are used to set them.

When establishing levels of proficiency in reading comprehension for students in various points in their school years, several kinds of performance standards can be defined: (1) adult standards, defined in terms of the terminal levels of textual difficulty that students need to be able to comprehend, (2) rational grade level standards, based on knowledge of the shape of the developmental learning curves for reaching terminal standards; or (3) de facto grade level standards, defined in terms of materials assigned to students. In the absence of information on adult standards or rational grade level standards, the expectations for reading comprehension, as defined by the Board of Regents*, were based on de facto grade level standards.

The materials assigned to students to be mastered during various points in their school years constitute a set of de facto standards for reading achievement. Thus these materials which students are required to read are the de facto standards for reading achievement.

---

*The Board of Regents is the governing body of the University of the State of New York and the State Education Department. The primary functions of the Board of Regents are the formulation of educational policy for the State and the exercise of powers which reside in the University. The policies and procedures established by the Regents are known as Regents Rules, and have the force and effect of law.

The DRP is uniquely suited to the competency program and the setting of standards. "Using the Degrees of Reading Power methodology it is possible to translate the difficulty of prose text into a statement of the ability required to process it. Conversely, it is possible to translate a statement of ability into a forecast of the difficulty of prose that can be understood. This means that it is possible for the DRP to be objectively criterion-referenced in terms of predicting performance on tasks external to itself. (Koslin, 1978c)." "The purposes of the DRP require that item difficulty be directly linked to passage difficulty. Standard item-writing techniques do not ensure that test items accurately reflect the difficulty of textual material. However, cloze procedures do generate items which are prima facie tied to the difficulty of the text in which they appear. Because DRP items are somewhat related psycholinguistically to cloze, and because of the rules of formation for the items, DRP items reflect the difficulty of the text in which they appear. (Koslin, 1978c, p. 37)."

Since "average" Rasch item difficulty per passage highly correlates with linguistic features of the text known to account for its difficulty and since "average" Rasch item difficulty per passage highly correlates with text readability, the DRP item difficulties have empirical meaning with respect to prose that students can understand. One regression equation allows DRP ability scores to be converted into forecasts of the prose that students can read. The other regression equation allows for the prediction of the ability required to comprehend prose of a given level of readability.

The regression equation predicting DRP difficulties from readability enables the DRP to be "objectively" criterion-referenced. Thus, if readability formulas are used to analyze a set of reading tasks, it is possible to use the results of these analyses to find the DRP passage difficulties which correspond in readability to minimal, average, and any other adult prose reading tasks. (Koslin, 1978a).

The correlation between the ordinal position of a passage on the DRP and its mean and median passage difficulty is essentially one. Since the DRP is constituted by writing passages at a pre-assigned level of readability using the Bormuth cloze formula, the correlation between Bormuth cloze scores and ordinal position must be one (Koslin, 1978a). Reference to Table 1 shows that this is the case.

Standards Expressed in Terms of the Difficulty of Text Which Needs to be Processed

The scaled difficulty (readability) of text assigned to students was used as the basis for defining standards of reading competency and as the basis for measuring progress toward the attainment of these standards. As Koslin (1973) noted, once the written materials that students of the educational system are expected to read are specified, and the materials scaled for readability, and the values appropriately aggregated, then the level of reading comprehension ability that students of the educational system must attain can be operationally defined by the "average" measured readability characteristics of the designated materials. Thus, performance standards can be established.

Once the quantative standards of reading competence are defined empirically, simple and direct measurements of the extent to which students have attained these standards can be made by administering the DRP. The test was designed to determine whether individuals can read with comprehension passages that have been scaled for the same linguistic properties that were used to define the standard, i.e. readability.

Sampling Plan

The sampling plan used to define de facto reading standards in grades five and six in New York State is an abbreviated version of the plan used by Carroll et al to calculate the 1969 national de facto readability standards in grades one through nine.

In order to calculate de facto reading standards in grades five and six in New York State it was necessary to determine which textbooks were most used in these grades. So as to be sure thatthe texts sampled would represent as nearly as possible the ranges of required reading to which fifth and sixth grade students are exposed, a five percent stratified

random sample of school districts was selected. The sample was stratified on the basis of community type[1] proportion of population within community type and geographic distribution.

A questionnaire was then mailed to the superintendents of the 41 selected school districts. The questionnaire was divided into five subject categories: English-Language Arts (reading), English-Language Arts (other), science, social studies, health and safety. Respondents were asked to list the textbooks commonly used in these subject areas in each grade and to specify for each text, the title, author, publisher and copyright date.

Selection of Textbooks for the Sample

Of the forty-one superintendents contacted, thirty-seven responded. Responses to the questionnaire were tabulated, and a grade by subject matrix was generated which indicated the total number of times that a book was listed for a particular subject and the number of different titles. If a book was mentioned for grade five and grade six it was entered in the matrix for each grade. This matrix served as the basis for the selection of texts to be sampled.

In order for a text to be sampled it had to be listed a minimum of three times or if not mentioned three times it had to be on of the texts selected by Carroll et al. The books selected were sampled in proportion to their pedagogical importance. That is the texts were weighted according to the time allotted for each curriculum area (Language Arts 46 .1%, social studies 30.8%, science 15.4% and health 7.7%).

---

1 Public and Nonpublic- The seven community type reference groups are: Type 1 - New York City; population over 1,000,000, Type 2 - Large Size Cities - Albany, Buffalo, Niagara Falls, Rochester, Syracuse, Utica, and Yonkers; population over 100,000, Type 3 - Medium Size Cities - Binghamton, Elmira, Mount Vernon, New Rochelle, Rome, Schenectady, Troy and White Plains; population 50,000 - 100,000, Type 4 - Small Size Cities - population under 50,000, Type 5 - Large Central Districts, (enrollment over 2,500 pupils and all Village Districts, Type 6 - Large Rural Districts, enrollment 1,100 - 2,500 pupils, Type 7 - Small Rural Districts, enrollment under 1,100 pupils.

Using the Samples of Textbooks to Define De Facto Reading Standards

The texts selected were accepted as a suitable set of texts with which to define the de facto reading standard for sixth grade students. Because of the way in which the texts were selected, the sample clearly represents those books in modal use in fifth and sixth grades in New York State.

Drawing Sample Passages from Textbooks to Define De Facto Standards

Once selected, textbooks were divided into sections depending upon the number of pages that they contained.

Materials containing at least two hundred pages were divided into tenths. Three passages of two hundred fifty words were randomly selected from the first, third, fifth, seventh, and final tenth of the book. Books containing one hundred to one hundred ninety nine pages were divided into fifths. From the first, third and final fifth of the text three passages of two hundred fifty words were randomly selected. Materials containing sixty to ninety nine pages were divided into thirds with the two hundred fifty word samples randomly selected from the first and last third of the book. Texts with less than sixty pages had three passages of two hundred fifty words randomly selected from the middle of the book.

The passages once selected were analyzed for readability by means of the Bormuth[2] readability formula. This formula takes into consideration such factors as the proportion of "easy" words (words on the Dale-Long list), word length (in number of letters), and sentence length. In addition, Bormuth adopted the Dale-Chall practice of counting the names of people, nationalities, places and certain other proper nouns as if they were common words.

---

2 The Bormuth formula is generally viewed as technically and linguistically superior to other formulas in the following respects: (a) it provides a continuous scale over the entire range of readability of first grade through at least twelfth grade instructional materials, (b) it has a higher validity coefficient in original development (r = .83) and in cross validation (r = .92) than other readability formulas, (c) it has relatively low standard errors of measurement (S_{em} = .06) in original development and (S_{em} = .04) in cross validation (d) cross validation data have been published (e) cloze test scores rather than multiple choice tests or expert judgements, were used to rate passage difficulty; cloze test scores are more objective than these other methods of rating passage difficulty.

The Bormuth formula was applied to each of the passages and the mean readability was then determined for each of the textbooks. Next the Bormuth readability index was converted to DRP units. DRP units = (1-B)100, where B equals the Bormuth value. DRP units form a scale of readability ranging from 15 to 100.

Results

The results of the analysis of the selected textbooks are summarized in Table 2 and Figure 1. In addition, the results of the texts analyzed by Carroll et al are summarized in Table 3.

From Table 2 it can be seen that the readability of textbooks most used is about the same for both grade five and grade six, the readability of textbooks is relatively homogeneous across subjects, and the readability of the textbooks is relatively homogeneous within subjects for grade six.

From Figure 1 it can be seen that textbooks become increasingly difficult over the school years. The reading demands, placed upon students, steadily increases.

Setting the Standard (State Reference Point)

Three possibilities existed upon which to base the reference point; (1) the mean readability in DRP units of fifth grade textbooks only ($\bar{x}$ = 54, σ = 5), (2) the mean readability in DRP units of sixth grade textbooks, only ($\bar{x}$ = 57, σ = 5), (3) the mean readability in DRP units of fifth and sixth grade textbooks combined ($\bar{x}$ = 55, σ = 5). The decision was made to base the standard upon the mean readability in DRP units of fifth and sixth grade textbooks combined. This was done because the test is administered at the beginning of sixth grade, and the average readability of fifth and sixth grade textbooks differ by only 3 DRP units. The standard error for the Bormuth readability formula is 4 DRP units.

Next it was necessary to determine the level of comprehension to be required of students. The obvious choices were; p = .65 the level of comprehension set for high school graduation; p = .75 the level of comprehension traditionally defined as the instructional level; or p = .70 a compromise between these two.

The State Reference Point was not set at p = .75 for fifth and sixth grade books as this would indicate that the student, at the beginning of grade 6, could already read fifth and sixth grade texts at the instructional level. The standard would be too difficult. The State Reference Point was not set at p = .65 the level for high school graduation. In setting this standard two types of errors can be made; (1) remediating students who do not need remediation and (2) failing to remediate students who need remediation. Since the Department is committed to the early identification of student reading problems, it was preferable, if an error was to be made, to remediate students who do not need remediation rather than fail to remediate students who need remediation. The standard was set at 55 DRP units at p = .70.

Conclusion

The textbooks analyzed were selected based upon the results of the questionnaire sent to the random sample of school districts and to a certain extent upon the Carroll survey. Information was also available on the readability of basal readers used in grades four, five and six by sixty one geographically representative school districts volunteering to field test the DRP. (Table 4). The textbooks analyzed were obviously representative of textbooks most used on a statewide as well as a national basis.

Chall's research (1979) has shown that textbooks with copyright dates in the 1960's tend to be more demanding than textbooks with later copyright dates. Therefore, if the sample of textbooks to be analyzed had been based only on the Carroll survey the required level of reading ability would have been more demanding than the level selected.

The reading demands, placed upon students steadily increases over the school years. The procedure of identifying students in need of remediation in grade 6 and grade 8 or 9 and setting standards based upon the readability of textbooks most used should be sufficient to carry out the Regents mandates of (1) assuring the early identification of students who need special help in developing their skills, and (2) assuring that students have acquired an adequate competence in these skills before graduation from high school.

A variety of questions arise from the setting of de facto standards one of which is: "Are the de facto standards high enough?" This question cannot be adequately answered until there is some systematic readability analysis of the reading tasks that adults actively engage in.

The question, however, can be addressed if the stand is taken that "the only sensible interpretation of data from assessment programs is based solely on whether the rate of performance goes up or down (Glass, 1978, p 259)." There is reason to believe that the level of reading performance will go up. As part of the Competency Testing Program, schools are required to provide special help to students who score below the State Reference Point on any state test. The reference point or standard was set with a view toward improving the level of reading statewide; to reverse the trend of declining scores on reading achievement tests. Toward this end four placement groups have been constituted in relation to the State Reference Point. Tables 5 and 6 list the four groups according to raw scores (total number right) on both forms of the sixth grade reading test. They also indicate the recommended action that should be taken with respect to each groups. The same information is provided for students taking the Preliminary Competency Test. In addition the DRP methodology aids teachers in selecting textbooks that match students reading ability. The State Education Department publishes readability reports that list commonly used textbooks and indicate their levels of difficulty or readability.

Forecasting Future Reading Requirements

For present purposes the New York State Education Department has not determined the readability of any systematic or representative sample of reading tasks that adults actually engage in or ought to be able to achieve. Such work lies in the future as part of the setting of rational standards.

The setting of rational standards will be based upon the reading requirements that citizens will need to meet fifteen to twenty years in the future. The program of reading instruction that a student receives from the time he enters school should be geared to preparing him for standards he will need to meet, at the time of high school graduation, in order to enter college, technical school, or an apprenticeship program to get a job, or to function effectively in the adult word in general.

To this end the Department is presently committed to performing readability analyses on adult materials. This commitment is based on the fact that adults must read more than they had to in earlier times. Each year there are fewer jobs that can be filled by persons with limited reading skills. These changes have resulted in a sharp increase in the reading demands on adults over the last generation. Therefore, to do an adequate job of educating students to be competent adult readers, it is necessary to be able to forecast the reading skills that a student entering first grade now will need by the time he is a graduating twelfth grader.

As additional information becomes available, the setting of standards for reading competency can be updated,

TABLE 1

Intercorrelations Between Mean and Median DRP Passage Difficulties and Bormuth Readability Values[a]

Form Y	Bormuth	Ordinal Position DRP Passage
Mean DRP Passage Difficulties	.980	.980
Median DRP Passage Difficulties	.971	.961
Bormuth Cloze Scores		1.000
Form X		
Mean DRP Passage Difficulties	.944	.944
Median DRP Passage Difficulties	.965	.965
Bormuth Cloze Scores		1.000

Note: [a]Mean cloze scores

NOTE: From "The Validity and Reliability of the Degrees of Reading Power Test" by B. L. Koslin, S. Koslin, H. Wainer, and S. Zeno. Touchstone Applied Science Associates, 1978. Reprinted by permission.

TABLE 2

MEAN READABILITY, STANDARD DEVIATION
AND NUMBER OF TEXTS ANALYZED
BY SUBJECT AND GRADE FROM
RANDOM SAMPLE OF NEW YORK STATE
SCHOOL DISTRICTS

Subject	Grade(s)	Number of Texts	Mean Readability in DRP Units
Language Arts	5	9	51
Language Arts	6	8	58
Social Studies	5	10	55
Social Studies	6	13	56
Science	5	2	54
Science	6	3	57
Health and Safety	5	3	54
Health and Safety	6	3	55
Language Arts	5 and 6	17	54
Social Studies	5 and 6	23	56
Science	5 and 6	5	56
Health and Safety	5 and 6	6	55
All Subjects	5	24	54
All Subjects	6	27	57
All Subjects	5 and 6	51	55

TABLE 3

MEAN READABILITY IN DRP UNITS
OF READING TEXTBOOKS
1969 - NATIONAL DE FACTO READABILITY
STANDARDS

GRADE	$\bar{X}$
1	37
2	43
3	47
4	49
5	52
6	53
7	54
8	52
9	55

TABLE 4

MEAN READABILITY IN DRP UNITS
OF READING TEXTS
USED BY NEW YORK STATE
SCHOOL DISTRICTS FIELD TESTING
THE DRP

Grade(s)	Number of Texts	Mean Readability in DRP Units
4	11	51
5	14	53
6	13	55
5 and 6	27	54

TABLE 5

PLACEMENT DECISIONS BASED ON THE READING TEST FOR NEW YORK STATE ELEMENTARY SCHOOLS FORM A1

Total Number Correct	Probability of Comprehending Textbooks Used in Fifth and Sixth Grade	Comment	Recommendation
82 or above	- - -	Form A1 inappropriate	Use Form A2
70 to 81	.70 and above	Above the State Reference Point	Increase demands of reading and writing tasks in the regular instructional program
67 to 69	.65 to .69	Below the State Reference Point	Provide remediation in the regular instructional program
59 to 66	.50 to .64	Far below the State Reference Point	Provide remediation beyond that included in the regular instructional program
58 or below	.49 and below	Very far below the State Reference Point	Provide intense individualized remediation

TABLE 6

PLACEMENT DECISIONS BASED ON THE READING TEST FOR NEW YORK STATE ELEMENTARY SCHOOLS FORM A2

Total Number Correct	Probability of Comprehending Textbooks Used in Fifth and Sixth Grade	Comment	Recommendation
44 or above	.70 or above	Above the State Reference Point	Increase demands of reading and writing tasks in the regular instructional program
41 to 43	.65 to .69	Below the State Reference Point	Provide remediation in the regular instructional program
33 to 40	.50 to .64	Far below the State Reference Point	Provide remediation beyond that included in the regular instructional program
32 or below	.49 and below	Very far below the State Reference Point	Provide intense individualized remediation

References

Bormuth, J.R. "Readability: A New Approach." Reading Research Quarterly, 1966, 1, 79-132.

Bormuth, J.R. "Cloze Test Readability: Criterion Reference Scores." Journal of Educational Measurement. 1968, 5, 189-196

Bormuth, J.R. Development of Readability Analyses. Final Report, Project No. 7-0052, Contract No. OEC-3-7-070052-0326. Office of Education, Bureau of Research, U.S. Department of Health, Education, and Welfare. March, 1969 (a).

Bormuth, J.R. "Factor Validity of Cloze Tests as Measures of Reading Comprehension Ability." Reading Research Quarterly. 1969, 4, 358-365. (b).

Bormuth, J.R. Development of Standards of Readability: Toward A Rational Criterion of Passage Performance. Final Report, Project No. 9-0237, Grant No. OEG-0-9-230237-4125, Bureau of Research, Office of Education, U.S. Department of Health, Education, and Welfare, 1971.

Byrne, C.D. "Setting Standards and Expectations for Reading Comprehension. Paper presented at AERA, San Francisco, California, April 10, 1979.

Carroll, J.B. Davies, P., & Richman, B. The American Heritage Word Frequency Book. Boston and New York; Houghton Mifflin Co. and American Heritage Publishing Co., 1971.

Chall, J.S. An Analysis of Textbooks in Relation to Declining SAT Scores. College Entrance Examination Board, 1977 (1-75).

Glass, G.V. "Standards and Criteria." Journal of Educational Measurement. 1978, 4, 237-261.

Koslin, S., Koslin, B.L., Langley, C., & Josephson, B. Design Concepts for A Measure of Effectiveness in Reading. New York, Riverside Research Institute, 1973.

Koslin, S., Koslin, B.L. "The Further Development of an Effectiveness Measure in Reading." New York, Touchstone Applied Science Associates, 1978. (a).

Koslin, S., Koslin, B.L., Wainer, H., Zeno, S. "The Validity and Reliability of the Degrees of Reading Power Test." Touchstone Applied Science Associates, 1978. (b).

Koslin , S., Koslin, B.L., Zeno, S. "The Further Development of an Effectiveness Measure in Reading: Phase III. The New York State Education Department. August, 1978c.

Linn, R.L. "Demands, Cautions and Suggestions for Setting Standards" Journal of Educational Measurement. 1978, 4, 301-308.

Meskauska, J.A. "Evaluation Models for Criterion-Referenced Testing: Views Regarding Mastery and Standard-Setting. Review of Educational Research, 1976, 46, 133-158.

Millman, J. Passing Scores and Test Lengths for Domain-Referenced Measures. Review of Educational Research, 1973, 43, 205-216.